DOCUMENT RESUME

ED 034 679 SE 007 562

AUTHOR Mowery, Kay A.
TITLE Information Requirements Analysis in a Secondary School System.
INSTITUTION Georgia Inst. of Tech., Atlanta. School of Information Sciences.
SPONS AGENCY National Science Foundation, Washington, D.C.
REPORT NO GITIS-69-13
PUB DATE 69
NOTE 44p.

EDRS PRICE EDRS Price MF-$0.25 HC-$2.30
DESCRIPTORS Administrative Problems, *Information Science, *Information Systems, Information Utilization, School Administration, Secondary Schools, *Systems Approach

ABSTRACT

Included are an examination of the formal structure of information collection and recording activities, and a discussion of the analysis of the resulting data. Although specific application of these ideas is applied to a secondary school system, the applicability of the formal structures and analytic methods is thought to be general. In section I, traditional structures are formalized: the school system's organizational structure and the functions, inputs, and outputs of the typical department. In section II, methods of analysis for functions and data are explored. In the final section, general procedures are suggested for the collection and recording of requirements and for analysis. Possible utilization of the computer in manipulating the large volumes of data is explored. The appendix contains a brief review of the literature on collection and analysis methods. (RR)

INFORMATION REQUIREMENTS ANALYSIS IN A SECONDARY SCHOOL SYSTEM

Kay A. Mowery

Georgia Institute of Technology

School of Information Science

Work reported in this publication was performed at the School of Information Science, Georgia Institute of Technology. The primary support of this work came from the Georgia Institute of Technology and from agencies acknowledged in the report.

* * *

The graduate School of Information Science of the Georgia Institute of Technology offers comprehensive programs of education, research and service in the information, computer and systems sciences. As part of its research activities the School operates, under a grant from the National Science Foundation, an interdisciplinary science information research center. Correspondence concerning the programs and activities of the School may be addressed to Director, School of Information Science, Georgia Institute of Technology, Atlanta, Georgia 30332.

Telephone: (404) 873-4211

GITIS-69-13

(Research Report)

INFORMATION REQUIREMENTS ANALYSIS IN A SECONDARY SCHOOL SYSTEM

Kay A. Mowery

School of Information Science 1969

GEORGIA INSTITUTE OF TECHNOLOGY

Atlanta, Georgia

ACKNOWLEDGEMENT

The work reported in the paper has been sponsored in part by the National Science Foundation Grant GN-655. This assistance is gratefully acknowledged.

VLADIMIR SLAMECKA
Project Director

ABSTRACT

This report provides an extended view of the formal information structures required in the requirements collection and recording phase of development. Analytic methods are suggested for use upon the resultant data as guidance to implementation. The ideas in the paper are applied to a secondary school system by way of example but the applicability of the formal structures and analytic methods derived is thought to be general.

INTRODUCTION

As the scope of information systems development has broadened to encompass the "total" systems concept, the mere extension of the systems-and-procedures methods of identification of input, processes and output has not been completely adequate to the task of systems design. Traditional methods have not contained information structures that reflect the nature of the total system, which may become a quite complex network of subsystems. In addition, traditional methods for single-function systems have not had to deal with the complex problem of deciding what to do first; that is to say, deciding which functional subsystem to implement first, or what systems-level data groups to consider first for machine support. Truly analytic methods directed to this problem have been totally lacking, with dependence placed upon the intuitive advice of the computer specialist or upon political exigencies.

In view of this situation, the intent of the present report is to examine the formal structure of large information-requirements collection and recording activities, and to discuss the analysis of the resulting data. The specific application of these ideas to a secondary school system is occasioned by the availability of a recent requirements report for which the writer was partially responsible (6). Although that report has been used to provide the examples cited in this paper, it is the writer's view that formal structures and analytic methods are applicable to many types of organizations and information systems.

A developmental approach has been taken in the paper in exploring structures and methods. In Section I, traditional structures are formalized: first the organizational structure, and then the functions, inputs, and outputs of the typical department. Additional system-level structures and secondary structures needed for analysis are then defined. In Section II, methods of analysis for functions and data are explored. Section III, general procedures are suggested for the collection and recording of requirements and for analysis. Possible utilization of the computer in manipulating the large volumes of data is explored. Throughout the paper, the ideas are illustrated using a small segment of requirements data extracted from the referenced report. Finally, the appendix contains a brief look at the literature on collection and analysis methods.

The author wishes to express his appreciation for the cooperation and assistance provided by Mr. Tom McConnell of the Atlanta Public Schools. The author is also grateful for the review and criticism rendered during the project by Professor Pranas Zunde and Mr. Robert Siegmann.

SECTION I: STRUCTURE

1.1 The Functional/Organizational Structure

In our model, the school system will be viewed as a community of user-processors of information, with a primary structure based on existing organization. This is essentially a functional[1] structure with minor deviations.

We will denote the system S and departments D, so that we may write:

$$S = \{D_i\} \qquad (1)$$

Since each department may also be a set of subdepartments, we have:

$$D_i = \{D_i^j\} \qquad (2)$$

We may then rewrite (1) above as

$$S = \{D_i^j\} \qquad (3)$$

with {i} defining a partition of S according to lines of authority.

The D-structure presented in Table I is based upon the organization of the Atlanta Public Schools as of Summer 1968. The structure is intended only as a representation, and no analysis or criticism will be made of it.

Table I
User/Processor Notation

D_0	Superintendent of Schools (Immediate Staff)
D_1	Controller
D_2	Administrative Service
D_3	Instruction
D_4	School Plant Planning and Construction
D_5	Staff Personnel
D_6	Research and Development

1. To avoid confusion, the mathematical concept of function will be always referred to as a mapping. Function here refers to an organizational work task and the term "process" will be reserved for sequences of functions. This usage is common.

D_{11}	Area I Schools
D_{12}	Area II Schools
D_{13}	Area III Schools
D_{14}	Area IV Schools
D_{15}	Area V Schools

D_1^0	Controller (Immediate Staff)
D_1^1	General Fund Accounting
D_1^2	Cafeteria Accounting
D_1^3	School Accounting
D_1^4	Special Project Accounting
D_1^5	Payroll Accounting
D_1^6	Bond Accounting
D_1^7	Budget/Savings Bonds/Summer School Tuition
D_1^8	Information Processing

D_2^0	Administrative Services (Immediate Staff)
D_2^1	Food Services
D_2^2	Maintenance and Operations
D_2^3	Purchasing
D_2^4	Administrative Services (Misc.)
D_2^5	Detectives

Table I, cont.

D_3^0	Instruction (Immediate Staff)
D_3^1	Curriculum Development and Supervision
D_3^2	Athletics, Physical Education and Military
D_3^3	Instructional Resources
D_3^4	Pupil Services
D_3^5	Vocational Education and Technical Education
D_3^6	Project Organization
D_3^7	dit
D_4^0	School Plant Planning and Construction (Immediate Staff)
D_4^1	Plant Planning
D_4^2	Land
D_4^3	Construction
D_4^4	Equipment
D_5^0	Personnel (Immediate Staff)
D_5^1	Recruitment and Placement
D_5^2	Certificated Personnel Support Services
D_5^3	Non-certificated Personnel Support Services
D_6^0	Research and Development (Immediate Staff)
D_6^1	Project Organization
D_6^2	dit

D_{11}^{0} Area I Office (Immediate Staff)

D_{11}^{1} Elementary School

D_{11}^{n} High School

Each department D_i may be considered as a subsystem with its own information requirements, functions and information products. In our already established set-theoretic notation we can refer to a department as a triple:

$$D_i = (S_i, I_i, O_i)$$

where S_i is the set of functions performed by department D_i; that is:

$$S_i = \{p_i^k\} \subseteq P, \text{ the set of all functions of the school system.}$$

I_i is the set of input data groups (or data sets) required by D_i in the accomplishment of S, i.e.:

$$I_i = \{i_i^k\}$$

and similarly O_i is the set of information[2] outputs of department D_i:

$$O_i = \{o_i^k\}$$

Note that, in the case of functions, the union of all S_i in the system generates the set P, i.e.:

$$\bigcup_D (S_i) = P$$

This is not so straightforward for inputs and outputs of D_i. The sets of interest here are F(the set of unique data groups within the system) and R(the set of informational products of the system). The difficulty is highlighted by noting that an output of one D_i may be the input to another. We can however note that:

$$\bigcup_D \{I_i\} = F \text{ and } \{O_i\} \subseteq (R \cup F)$$

where the elements of I_i incorporate data groups which are either transactional data or file records, and where the elements of O_i may be either

2. Other departmental products or outputs, either physical or intangible, are not shown in this presentation.

terminal outputs (formal reports) or data sets of F.

Referring again to the requirements study (6), we can extract as examples the (S,I,O)—structures for two subdepartments. These are presented in Table II.

Table II.a.

Structure of D_1^5: Payroll Accounting Department

1. The S-Structure: Functions

$(p_1^5)_1$	Payroll records maintenance
$(p_1^5)_2$	Professional payroll preparation
$(p_1^5)_3$	Non-professional payroll preparation
$(p_1^5)_4$	Student payroll preparation
$(p_1^5)_5$	Payroll funds accounting
$(P_1^5)_6$	Tax and benefits reporting (periodic)

2. The I-Structure: Input data groups

$(i_1^5)_1$	New employee background data
$(i_1^5)_2$	Employee action transmittals
$(i_1^5)_3$	Time/attendance -- personnel
$(i_1^5)_4$	Time -- student
$(i_1^5)_5$	Employee employment history
$(i_1^5)_6$	Employee payroll history
$(i_1^5)_7$	Student payroll record

3. The O-Structure: Output data groups

$(o_1^5)_1$	Payroll checks
$(o_1^5)_2$	Payroll register

Table II.a. cont.

The O structure, cont.

$(o_1^5)_3$	Deductions register
$(o_1^5)_4$	Year-to-date register (salary and attendance)
$(o_1^5)_5$	Automatic voucher register
$(o_1^5)_6$	Hospitalization and medical register
$(o_1^5)_7$	Records corrections audit lists
$(o_1^5)_8$	Employee employment history
$(o_1^5)_9$	Employee payroll history
$(o_1^5)_{10}$	Tax forms
$(o_1^5)_{11}$	Tax reports
$(o_1^5)_{12}$	Employee savings bonds

Table II.b.

Structure of D_5^2: Certificated Personnel Support Services

1. The S-Structure: Functions

$(p_5^2)_1$	Career planning assistance
$(p_5^2)_2$	Payroll certification
$(p_5^2)_3$	Professional certification, grants, and prior service
$(p_5^2)_4$	State reimbursement claims processing
$(p_5^2)_5$	Fringe benefits assessment
$(p_5^2)_6$	Records and Statistics handling

2. The I-Structure: Input data groups

$(i_5^2)_1$	Teacher career development actions
$(i_5^2)_2$	Placement notices
$(i_5^2)_3$	Employee background record (new)

Table 11.b. cont.

The I structure, cont.

$(i_5^2)_4$	Personnel pos./pay/record changes
$(i_5^2)_5$	Special earnings reports
$(i_5^2)_6$	Employee employment history record
$(i_5^2)_7$	Employee payroll and position record
$(i_5^2)_8$	Grant applications
$(i_5^2)_9$	Grant allocation notices
$(i_5^2)_{10}$	Payroll step status
$(i_5^2)_{11}$	State fund allocation
$(i_5^2)_{12}$	Benefits schedules and descriptions
$(i_5^2)_{13}$	Personnel actions/communications

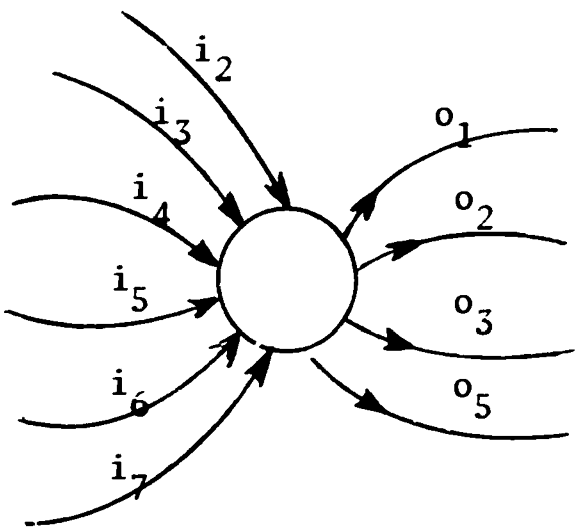
3. The O-structure: Output data groups

$(o_5^2)_1$	Employee payroll and position record (new)
$(o_5^2)_2$	Employee action transmittals
$(o_5^2)_3$	Employee notifications
$(o_5^2)_4$	Personnel classification and analysis report
$(o_5^2)_5$	Personnel salary report
$(o_5^2)_6$	Personnel certification report
$(o_5^2)_7$	State teacher requisition report
$(o_5^2)_8$	Personnel position lists
$(o_5^2)_9$	Automatic salary increment report

1.2 Analysis of Structure:

The structure which has been defined so far is the input-process-output inventory normally included in a systems analysis: all that is missing is the process flow chart linking the three and the system-level process-chart linking processes.

While it is not the primary concern of this paper, being well handled elsewhere (7,14), the intradepartmental structure (i.e., the internal structure of D_i^j) can be formalized by the use of mathematical relations and depicted in terms of matrices or graphs. For example, a portion of department D_5^2 could be viewed as a directed graph:



and relations of input to processes reflected by an incidence matrix (where a "1" indicates that i_k is an input to p_i and a "0" indicates there is no connection):

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}	i_{11}	i_{12}	i_{13}
p_1	1	0	0	0	0	1	0	0	0	0	0	0	0
p_2	0	1	1	1	1	1	1	0	0	0	0	0	0
p_3	0	0	1	0	0	1	0	1	1	1	0	0	0
p_4	0	0	0	0	0	0	1	0	0	0	1	0	0
p_5	0	0	0	0	0	0	1	0	0	0	0	1	0
p_6	1	1	1	0	0	1	0	1	0	0	0	0	1

Process-to-output relations can be shown by a similar matrix. A corresponding set-theoretic notation would take the form of ordered triples:

$$D_5^2 = \{(p_1, i_1, \phi), (p_2, i_2, o_1), (p_2, i_2, o_2), \ldots\ldots\} ,$$

an enumerative process. These examples are presented merely to show some of the formalisms now used to analyze system structure. To pursue this vein further, see one or more of the references noted above.

Utilizing one or more of these notational techniques with matrix algebra, we are able to build substantial models of the department, of the functional sector of the school system, and of the complete system. There are techniques to reduce complexity, and meaningful analysis of the system can be performed using these representations.

The structural preliminaries have the effect of forcing an exhaustive enumeration of the classes S, I, and O for each department. The immediate difficulties in simply generating department structures indicate the incompleteness of data in the standard requirements report. Two factors must be considered:

1. Too great a depth of detail must be avoided, else the manpower expenditure in collecting data is prohibitive; and

2. Change becomes a detiorating factor as the depth of structuring is increased.

Practical considerations work against a hard structuring of the type suggested. For one thing, it is difficult to get identical functions, inputs or reports so identified and named. If they are separated physically, there is an inertia based on past usage and an intradepartmental bias to maintain separate names; this works against any effort to mechanize even the purely structural aspects of analysis. The attempt in the subject study (6) to enumerate functional overlaps is something that might be expected of a structural analysis, but on the basis of the data as available and presented, it would be impossible to make a serious effort at such analysis. The inclusion of such data in the report was evidence of mental assimilation on the part of the project staff, rather than of a formal process.

Intra- and interdepartmental structures, however, are not the primary interest of this paper. We want to look at the sets P, F, and R, and to concern ourselves with structures which may be imposed on these sets during the requirements definition. Through the use of such structures, we hope to be able to make concrete recommendations about subsequent information system development.

1.3 Structures of P, F, and R

While the organizational (i.e., departmental) structure of the school system provides the most apparent and predominant ordering of the sets P, F and R, it is also the strongest deterrent to our ability to view the complete system. On each of the sets, therefore, we want to explore other ordering relations which might provide insight into the desired behavior of the school system as a whole for guidance in organizing processes and data for an automated environment. In this section we will postulate a variety of secondary structures, keeping in mind the need to further analyze and organize P, F, and R. We will assume that the definition of the departmental structure is complete, giving us an enumeration of the elements of P, F, and R as defined in Section 1.1.

1.3.1 Structures on the Set of Functions P

We have already noted the most prominent structure on the set of functions — the organizational structure of Section 1.1. This we can note formally as a mapping:

R_{PD}: $P \rightarrow D$ where the mapping relates a specific function to an organization. For example,

$$R_{PD}(p_5^2)_1 \rightarrow D_5^2$$

Of course this is *a posteriori* according to our previous definition; the new notation provides no further enlightenment as to the composition of P.

We would like to extract one or more additional structures of the set of functions from the data supplied in the requirements. Three of

these are identifiable and will be called program importance, organizational importance, and precedence.

1) Program importance, R_{PG}, is a mapping on the set of functions which associates each function with an identified program within the school system. Formally:

$$R_{PG}: \; P \rightarrow G$$

where G is the set of formal goal-oriented programs of the school system. The set G will be derived in Section 1.4. Because each of the members of G assumes a relative priority within the set, this priority carries over to, or is induced on, the elements of P which are mapped into G.

2) Organizational importance, R_{PM}, is a mapping of the set of functions which associates each function with one of the traditional activities within which the organization operates. Formally:

$$R_{PM}: \; P \rightarrow M$$

where M is a set of activities based on traditional line-staff relationships. We will derive the elements of M in Section 1.4 and indicate an order on the set which, again, we will want to induce on P.

3) Precedence, R_{PW}, is a chronological ordering of functions based on inherent cycles of operations within the organization. It defines a set of processes into which each function can be mapped. Formally:

$$R_{PW}: \; P \rightarrow W$$

Within W we have many processes, and each process w_k is strictly ordered internally. This internal ordering of processes may be used to provide guidance as to the order in which the elements of P are to be developed. The set W is partially connected and does not induce a partition as do the other two structures; however, we will see that W still proves useful in analysis.

1.3.2 Structures on the Set of Data F

As was the case with the set of functions, the most obvious structuring of data occurs as a result of the dominant departmental structure. We could also consider a structuring based upon program or activity, but such mappings are only indirectly useful to an analysis of the set F. Looking for additional structures on F we find the intuitive structure based on the object for which the data set is an attribute. In addition, we find that the sets D and G each generate a measure which provides an important descriptive structure on F.

1) Object attribute, R_{FC}, is a mapping of the data sets which corresponds to the intuitive structuring of data according to the real-world object it is describing. Formally:

$$R_{FC}: \ F \rightarrow C$$

where C is the set of object classes. The elements of C which are pertinent in the school system are defined in Section 1.4. The set C is the primary partition of data for machine applications.

2) Department use, R_{FU}, is a mapping of the data sets which is generated by the department structure. It maps each data set onto a measure of use or demand U. Formally:

$$R_{FU}: \ F \rightarrow U$$

where U is a discrete measure in the positive integers. It might be termed a demand index, since it reflects the number of departments utilizing a specific data set. It will be briefly discussed in Section 1.4.

3) Program use, R_{FV}, is a mapping similar to departmental use in that the set of programs generates a mapping of the data sets onto a value structure or measure V. Formally:

$$R_{FV}: \ F \rightarrow V$$

where V is a measure over the positive real numbers. Again it is an index, associating with each data set a number reflecting the composite importance of the programs by which it is used. Its construction will be briefly discussed in Section 1.4.

Later in this paper we will present methods of generating these structures on F, and we will attempt to consolidate the three; our intent is to postulate an organization on F and a priority for implementation.

1.3.3 Structures on the Set of Information Products R

The set R provides us with information on the organization; for instance, a partial structuring of R tells us which departments are information generators and which are only users. Thus, we have to look elsewhere for internal order and structure for R. Three structures can be postulated, two (urgency and primacy) based on use, and one (source originality) based on function.

1) Urgency, R_{RT}. Probably the most readily obtained attribute of a product is the frequency with which it is produced. While this tends to be predicated in terms of existing capabilities rather than of actual urgency, it does give us a measure of immediacy in terms of use. This attribute then can be described as:

$$R_{RT}: R \rightarrow T$$

where T is a set of time intervals of increasing size. A complete enumeration will appear in Section 1.4.

2) Primacy (or importance), R_{RX}. Each informational product is related to one or more functional responsibilities of the system as reflected by the activity structure M. Primacy, then, is a mapping of the products onto an importance measure. Formally:

$$R_{RX}: R \rightarrow X$$

where X is a measure defined on the positive real numbers. The method of generation for X will be discussed in Section 1.4.

3) Source originality, R_{RN}. This is essentially a dichotomous partitioning of information products into those which are generated and increase or change F and those which are merely a regeneration and restatement based on F. The generated members of R usually result from data collections as a part of the function, and the regenerated members are usually thought of as terminal products. The mapping is formally:

$$R_{RN}: R \rightarrow N$$

where N is a two-element set: source data and processed data. It is occasionally convenient to consider some further division of processed data to better describe the output report, e.g.: analysis, summary tabulation, detail tabulation, projection, etc. However, our main reason for declaring the set N is to reflect that retained products do exist and are within our definition of data. We will make little or no use of terminal products.

1.4 Summary of Structures With Examples

The previous section identified several subordinate structures (or at least sublimated structures) which tend not to be brought forward in the analysis of requirements. It might be well now to summarize the content of that section and to present examples of structures with which to work in later sections. Such a summary is found in Table III.

Table III

Mapping	Domain	Range
R_{PD} : organization	P	D : departments
R_{PG} : program importance	P	G : programs
R_{PM} : org. importance	P	M : states of activity
R_{PW} : precedence	P	W : processes
R_{FC} : object attribute	F	C : object classes
R_{FV} : program use	F	V : value index
R_{FU} : department use	F	U : demand index
R_{RT} : urgency	R	T : time intervals
R_{RX} : primacy	R	X : importance index
R_{RN} : source/orig.	R	N : source or proc.

It is critical to requirements analysis that both set G and set M be formally declared and that the priorities within each be agreed upon. Table IV will enumerate the programs and activities and establish ranking within the list. It should be noted that there can be no real incompatibility between the two sets, although some inconsistencies may show up when related to the department organization.[3] We will make multiple use of both G and M in structuring our system sets P, F and R.

3. These were generated from discussions with Mr. Tom McConnell of the Atlanta Public Schools, but are intuitive and informal in nature, based on personal judgments. An attempt at formal definition is under way but in no way affects the general use as here proposed.

Table IV

	G			M	
No.	Element Name	Prior.	No.	Element Name	Prior.
g_1	management	1	m_1	oper'l instruct.	1
g_2	oper. services	2	m_2	oper'l services	2
g_3	pupil services	3	m_3	oper'l admin.	3
g_4	normal instruct.	4	m_4	mgmt. control	4
g_5	athletics	5	m_5	mgmt. admin.	5
g_6	spec. instruct.	6	m_6	mgmt. planning	6
g_7	program planning	7	m_7	staff services	7
g_8	research/devel.	8	m_8	staff instruct.	8
			m_9	staff admin.	9

The object classes C are tentatively drawn within the requirements document without any real recognition of the process by which they were selected. That this is a very natural classification of data is admitted, but here we wish to call attention to it emphatically. With no regard to internal precedences, the set of classes of objects within the school system are shown in Table V.

Table V

The Elements of C

c_1	:	pupil data
c_2	.	personnel data
c_3	:	financial data (incl. funding, budgeting, acctg.)
c_4	:	facilities data (incl. installed equip., land)
c_5	:	equipment data
c_6	:	instructional resources
c_7	:	system structure data[4]

The set of processes W is best defined in certain specialized applications, such as process control; in general data processing, it tends to be well defined only in fragments, and then only after key decisions are made. One advantage of mechanical processing of function statements is the sorting out of processes. For our purposes here, we will indicate precedences among the elements of P as they exist in our examples in Section 1.1. Table VI enumerates the set W to the extent possible from our examples.

In the structures we have chosen as our examples, we find sequences occurring only in w_1, w_2, w_3, w_4 and w_7. If we were to add a school, or the Personnel Recruiting Department, for example, we would see more complete sequences, and the enumer_tion would be more complex.

4. The latter set constitutes an external system data class which must be available and includes organization data, staffing structure, etc.

Table VI

The Elements of W

w_1	:	$\ldots(p_5^2)_2,\ (p_1^5)_1,\ \ldots$
w_2	:	$\ldots(p_1^5)_1,\ (p_1^5)_2,\ (p_1^5)_5,\ldots$
w_3	:	$\ldots(p_1^5)_1,\ (p_1^5)_3,\ (p_1^5)_5,\ldots$
w_4	:	$\ldots(p_1^5)_4,\ (p_1^5)_5,\ldots$
w_5	:	$\ldots(p_1^5)_6,\ldots$
w_6	:	$\ldots(p_5^2)_1,\ldots$
w_7	:	$\ldots(p_5^2)_3,\ (p_1^5)_1,\ldots$
w_8	:	$\ldots(p_5^2)_4,\ldots$
w_9	:	$\ldots(p_5^2)_5,\ldots$
w_{10}	:	$\ldots(p_5^2)_6,\ldots$

The sets V, U, and X are defined without difficulty. The set U is composed of positive integers, where u_i is the number of departments using a data set; similarly, the set V is a set of positive reals reflecting again multiplicity of use (this time by programs), but also incorporating the priorities. X is a set of positive reals reflecting multiplicity of use for products, again making use of the priority structure of M. Since the computations are straightforward, we will postpone examples until they are needed.

The time intervals T are readily derivable from the requirements document, since they are among the concrete elements of reports descriptions. In Table VII we generate those time intervals found in Section 4.8 of the subject study (6). Finally, in Table VIII we list the set N for completeness.

Table VII

The Elements of T

t_1	:	daily
t_2	:	weekly
t_3	:	monthly
t_4	:	20 days
t_5	:	quarterly
t_6	:	annually
t_7	:	process cycle
t_8	:	end of project (length of project)
t_9	:	on demand/as required: immediate
t_{10}	:	on demand/as required: elapsed time response

Table VIII

The Elements of N

n_1	:	source data
n_2	:	processed data

SECTION II: ANALYSIS OF REQUIREMENTS

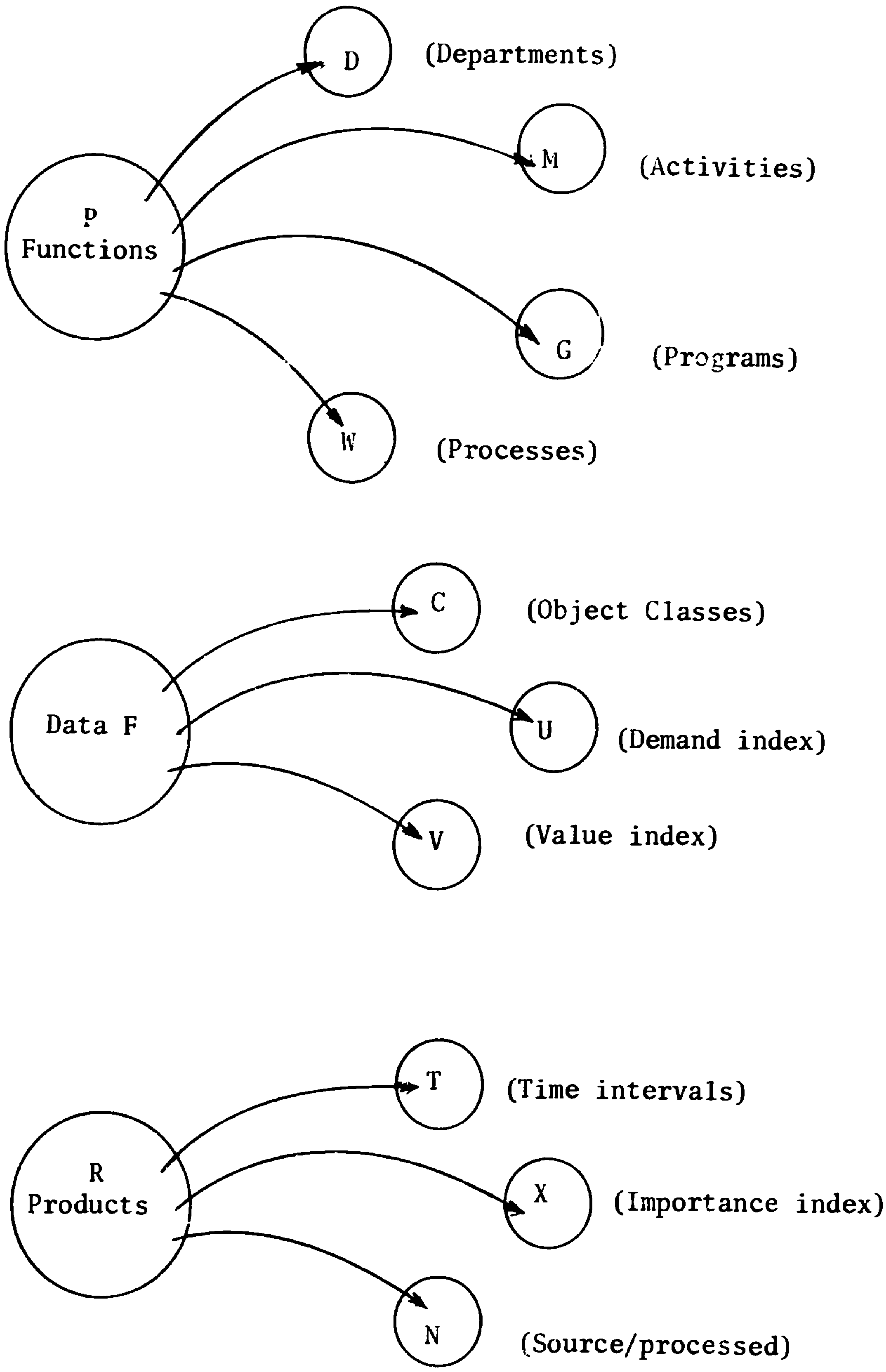
2.1 Introduction

In the first section we presented a reasonably complete catalog of structures within the school system. Out of the primary system components (functions, P; data, F; and products, R) we obtain the raw resources for our information system. The secondary structures (see Table III) represent a variety of influences on the primary components; and, if viewed in some coordinated fashion, can provide us with considerable insight into design approach. A simple diagram of the relationships among structures may be helpful before proceeding; see Figure I.

As indicated in Figure I, each element in P, F, and R has associated with it attributes induced by the mappings shown. This means that each element of P, for instance, has acquired as an attribute an element each from D, M, G, and W. Each of these attributes makes its demands on p_i, imposing on P a complex structure. If we understand this structure then we will be better able to make the proper decisions concerning implementation phasing.

In this section we will investigate systematic techniques for deriving the induced structures of P and F, and we will inquire whether or not guidance can be obtained which is at least consistent with current intuitive practice.

Figure I


D
(Departments)
M
(Activities)
P
Functions
G
(Programs)
W
(Processes)
C
(Object Classes)
Data F
U
(Demand index)
V
(Value index)
T
(Time intervals)
R
Products
X
(Importance index)
N
(Source/processed)

2.2 Analysis of the Set P

Taking the D-structures defined in Section 1.1 for the Payroll Accounting Department and the Certificated Personnel Support Services Department, our first requirement is to generate the mappings of P onto M, G, and W. This is shown in Table IX.

Table IX

Function (p_i)	Program (g_i)	Activity (m_i)	Process (w_i)
$(p_1^5)_1$	g_2	m_7	w_1, w_2, w_3, w_7
$(p_1^5)_2$	g_2	m_7	w_2
$(p_1^5)_3$	g_2	m_7	w_3
$(p_1^5)_4$	g_2	m_7	w_4
$(p_1^5)_5$	g_2	m_4	w_2, w_3, w_4
$(p_1^5)_6$	g_2	m_7	w_5
$(p_5^2)_1$	g_2	m_7	w_6
$(p_5^2)_2$	g_2	m_4	w_1
$(p_5^2)_3$	g_2	m_9	w_7
$(p_5^2)_4$	g_2	m_9	w_8
$(p_5^2)_5$	g_7	m_6	w_9
$(p_5^2)_6$	g_2	m_7	w_{10}

The partitioning of P according to D is reflected in the indexing of the p_i^j elements directly. By inspection we see that both G and M generate (induce) partitions on P.

In our example[5]:

$$P/G \; : \; \{[g_2],[g_7]\}$$

$$: \; \{\{(p_1^5)_1,\ldots,(p_1^5)_6,(p_5^2)_1,\ldots,(p_5^2)_4,(p_5^2)_6\},\{(p_5^2)_5\}\}$$

$$\text{and } P/M \; : \; \{[m_7],\ [m_4],\ [m_6],\ [m_9]\}$$

$$: \; \{\{(p_1^5)_1,\ldots,(p_1^5)_4,(p_1^5)_6,(p_5^2)_6\}\ ,\ \{(p_1^5)_5\}\ ,$$

$$\{(p_5^2)_2,\ (p_5^2)_3,\ (p_5^2)_4\},\ \{(p_5^2)_5\}\}.$$

In the case of W, however, the mapping is ambiguous in some cases, failing to produce clean partitions. The w_i, however are well defined, and we can enumerate their contents in order; this was actually done during the generation of Table VI, Section 1.4. The w_i's there are partially defined. ordered sequences. The structure generates both precedences and connectivity for the p_i^j involved. Thus:

$$P/W : \{\{\ldots(p_5^2)_2,(p_1^5)_1,\ldots\},\{..(p_1^5)_1,(p_1^5)_2,(p_1^5)_5..\},$$

$$\{..(p_1^5)_1,(p_1^5)_3(p_1^5)_5,\ldots\},\{..(p_1^5)_4,(p_1^5)_5,\ldots\},$$

$$\{..(p_1^5)_6,\ldots\},\{..(p_5^2)_1,\ldots\},\{..(p_5^2)_3,(p_1^5)_1,\ldots\},$$

$$\{..(p_5^2)_4,\ldots\},\{..(p_5^2)_5,\ldots\},\{..(p_5^2)_6,\ldots\}\}$$

We need to establish a precedence order between G and M and will therefore state that G has priority (precedence) over M, or:

$$G > P$$

This is reasonable, since G most closely approximates a goals-structure on the system.

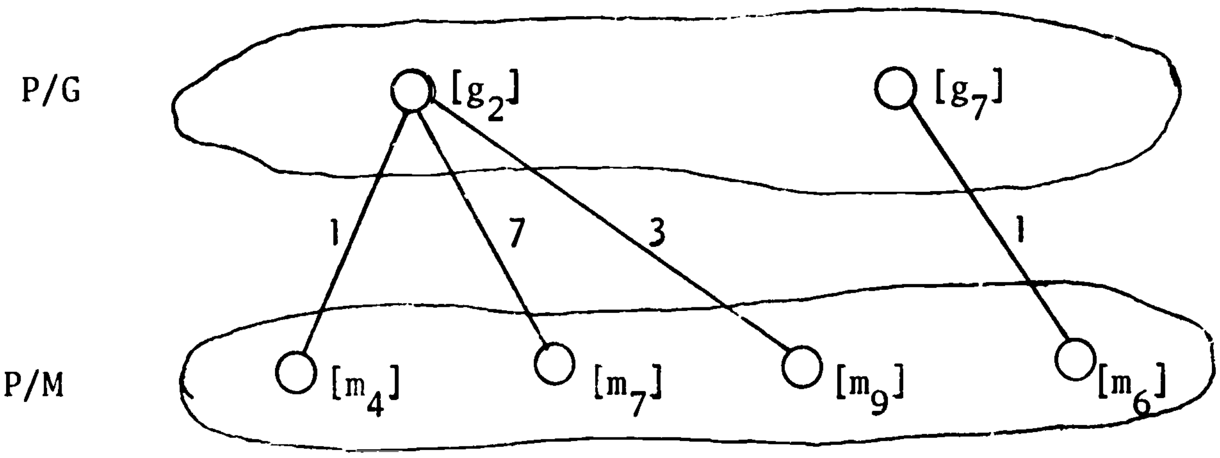
Our initial structure of P corresponds to P/G above. We now want to modify this structure using P/M. This we can do by considering a graph-theoretic representation of the partitions relative to both G and M (see Figure II).

---

5. We assume for convenience the notation of algebraic quotient structures to reflect the partitioning of our primary sets P, F, and R. Hence:

P/G denotes the partitioning of P induced by G and $[g_i] \subset P/G$ indicates a partition or subset of P of all functions having like program mapping.

Figure II



Here the graph indicates a partitioning of $[g_2]$ by M, with the numbers on the edges indicating the number of elements of P in each partition. Using our predetermined priorities within G and M (from Section 1.4) and the precedence between G and M we have an immediate set of priorities:

Pri 1: $[g_2]$
Pri 2: $[g_7]$

Since $[g_7]$ has only a single member, $(p_5^2)_5$, we would like to further subdivide Priority 1. Here, an ambivalence develops. We may either take the priorities of M as our secondary order, resulting in:

H1: Pri 1: $[g_2]/m_4$
Pri 2: $[g_2]/m_7$
Pri 3: $[g_2]/m_9$
Pri 4: $[g_7]$ m_6

or we may consider the size of the subsets developed, i.e.,

H2: Pri 1: $[g_2]/m_7$ (7 members)
Pri 2: $[g_2]/m_9$ (3 members)
Pri 3: $[g_2]/m_4$ (1 member)
Pri 4: $[g_7]/m_6$ (1 member)

It is at this point that we utilize the remaining structure on P, the set W. Since the elements of W (i.e., processes) are ordered sequences

of p_i^j which cut across G and M, we may find a solution. If we look at the functions p_i^j contained in g_2 we get a graph of the type shown in Figure III. If we delete the nodes (functions) which are isolated, we get an order graph such as that in Figure IV.

Figure III

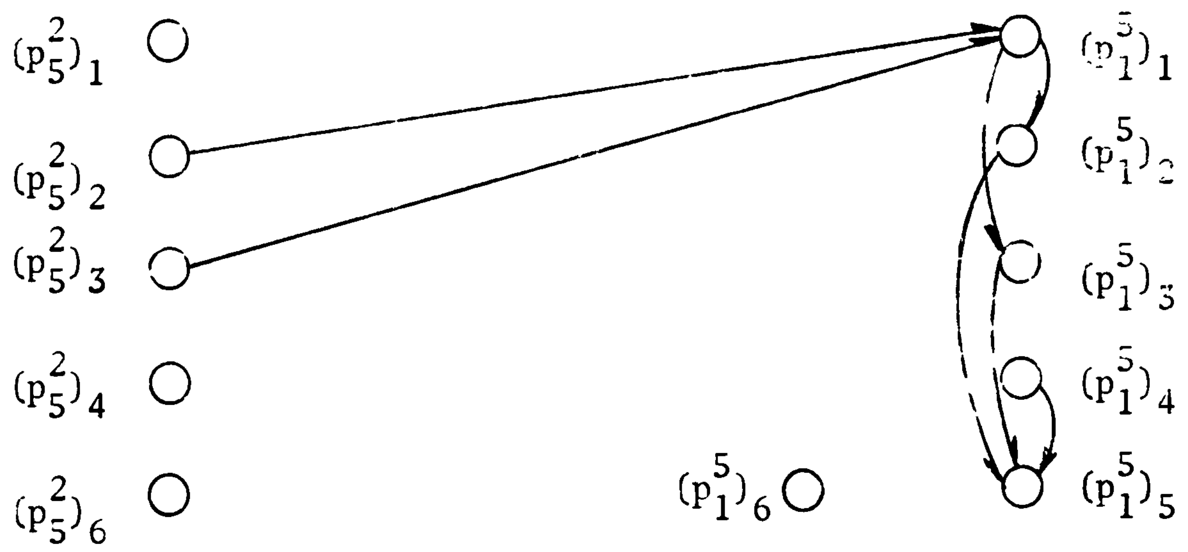
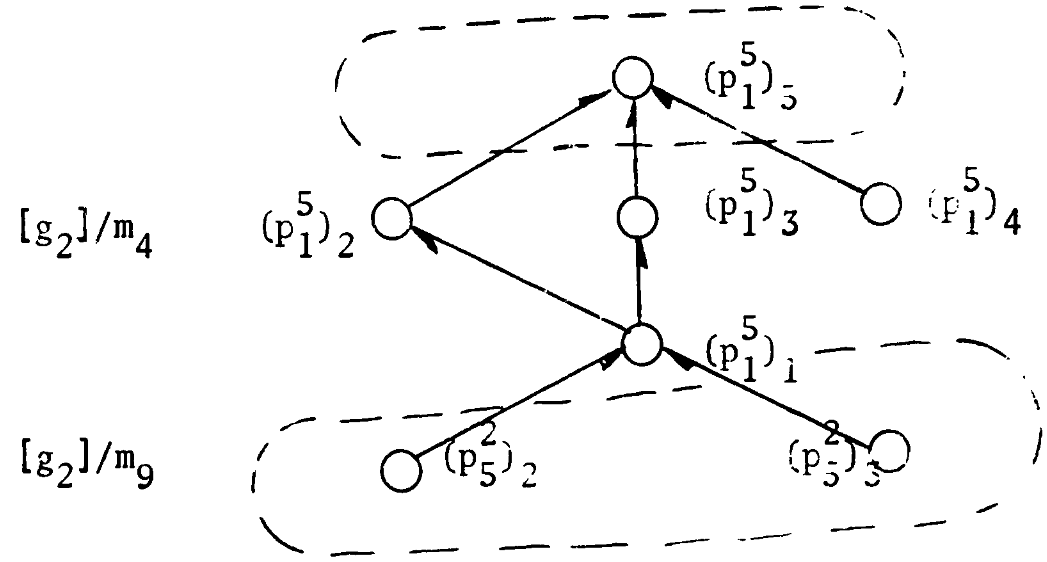


Figure IV



Inspection of this graph provides us with a set of priorities which are intuitively attractive:

Pri 1: $[g_2]/m_4$

Pri 2: $[g_2]/m_7$

Pri 3: $[g_2]/m_4$

Pri 4: $[g_7]/m_6$

In order to accompiish the $[g_2]/m_4$ function (fund accounting), we have to generate the necessary data through the preceding functions which make up the pertinent processes within the financial system. This, then, is a natural sequence for implementation.

2.3 Analysis of System Data: the Set F

Civen a complete and exhaustive definition of the system in terms of the department structure (S,I,O), we are in a position to generate the system data set F. This is essentially a tabulation process, in which duplicate use of a data group, i_i^j, is recorded, generating for us the corresponding use index $u_i \in U$ and the program-value index $v_i \in V$ according to its required use within the system. This is convenient, since we have previously made the linkage of program to function. We will use only the I-structure to generate the set F, i.e.:

$$F = \bigcup_D I_i$$

All nonterminal outputs, those which are source data sets to be reused by the system, will be picked up in the I-inventory.

In generating the set F, we wish also to establish the object class associated with each data set through the R_{FC} mapping. While the U- and V-generation was mechanical in nature, this assignment of object class is analytical and to a degree arbitrary. Table X presents a tabulation of data sets from our example structures (Section 1.1). It was necessary to extrapolate to obtain the U and V values, but the accuracy is adequate for our purposes.

Table X

Element (f_i)	Description	Cross/Ref. (Sect. 1.1)	Object Class	User[6] Demand	Value[7] (v_i)
f_1	new emp bkgrd	$(i_1^5)_1$, $(i_5^2)_2$	c_2	4	2.5
f_2	empl action	$(i_1^5)_2$	c_2	4	3.0
f_3	time/att-pers	$(i_1^5)_3$	c_2	3	2.0
f_4	time-student	$(i_1^5)_4$	c_1	3	2.0
f_5	emp history	$(i_1^5)_5$, $(i_5^2)_6$	c_2	10	3.9
f_6	emp pay hist	$(i_1^5)_6$, $(i_5^2)_7$	c_2	7	3.4
f_7	stud pay hist	$(i_1^5)_7$	c_1	3	2.0
f_8	career dev ac	$(i_5^2)_1$	c_2	4	3.0
f_9	place notice	$(i_5^2)_2$	c_2	4	3.0
f_{10}	emp p/p/r chg	$(i_5^2)_4$	c_2	5	3.5
f_{11}	emp spec pay	$(i_5^2)_5$	c_2	4	3.0
f_{12}	emp grant apl	$(i_5^2)_8$	c_2	3	2.5
f_{13}	emp grant alc	$(i_5^2)_9$	c_3	5	3.5
f_{14}	pay step stat	$(i_5^2)_{10}$	c_3	5	3.5
f_{15}	state fund al	$(i_5^2)_{11}$	c_3	5	3.5
f_{16}	bene sched/de	$(i_5^2)_{12}$	c_7	5	3.1
f_{17}	pers commun	$(i_5^2)_{13}$	c_2	3	2.0

6. These numbers were generated from the requirements study based on the writer's participation in the collection cycle and utilize the data matrix in Section 4.8 of the study. They indicate the number of second-level departments having the requirement.

7. Using the departments reflected in user demand, the program priority related to each function was generated. A composite rank was then computed using the following formula:

$$V_i = \sum_D \frac{1}{G(D_i)} \text{ for } D_i^j \text{ included in } u_i$$

2.3.1 Partitioning of Data by Object Class

The most natural subdivision of the set F is according to the related object classes. From the table we see the following partitioning:

$$F/C : \{[c_1], [c_2], [c_3], [c_7]\}$$

where $[c_1] = \{f_4, f_7, ..\}$: pupil data

$[c_2] = \{f_1, f_2, f_3, f_5, f_6, f_8, f_9, f_{10}, f_{11}, f_{12}, f_{17}, ..\}$: personnel data

$[c_3] = \{f_{13}, f_{14}, f_{15}, ..\}$: financial data

$[c_7] = \{f_{16}, ..\}$: system data.

Based on the data itself, we cannot legitimately say anything about the relative importance of object classes. Each class is necessary to the functioning of the school system; only within subsystems might such comparisons be made. We are in a position, however, to say something about the relative importance of data sets within an object class. To illustrate, we will use the U and V mappings identified in Section I; the data sets from our two example departments; and the Table X estimates. Only $[c_2]$, personnel data, has a significant number of data elements included, but we will briefly look at all four classes.

2.3.2 Importance Within Object Class

Looking first at $[c_1]$, pupil data, we see that successive application of partitioning by U and V produces no results:

$$[c_1]/U = [c_1] = \{f_4, f_7\}$$

$$[c_1]/U/V = [c_1] = \{f_4, f_7\}$$

No order of importance can be derived, since both f_4 and f_7 have a U-index of 3.0 and a V-index of 2.0. This is due to the lack of a significant portion of $[c_1]$ available. The relative importance in the table indicates, however, that these two data sets are probably not of major importance within their class.

Looking next at $[c_2]$, personnel data, we get a partitioning of the

data sets which produces a spectrum of importance:

$$[c_2]/U = \{[u_3], [u_4], [u_5], [u_7], [u_{10}]\}$$

which by order of importance becomes:

I1: $[u_{10}]$: f_5

I2: $[u_7]$: f_6

I3: $[u_5]$: f_{10}

I4: $[u_4]$: $f_1, f_2, f_8, f_9, f_{11}$

I5: $[u_3]$: f_3, f_{12}, f_{17}

If we further order $[u_3]$ using V, we get:

$$[u_3]/V = \{[v_{2.0}], [v_{2.5}]\}$$

where $[v_{2.0}]$: $\{f_3, f_{17}\}$

$[v_{2.5}]$: $\{f_{12}\}$

If we further order $[u_4]$ on V, we get:

$$[u_4]/V = [v_{2.5}], [v_{3.0}]$$

where $[v_{2.5}]$: $\{f_1\}$

$[v_{3.0}]$: $\{f_2, f_8, f_9, f_{11}\}$

Incorporating this substructure derived from V, we obtain the following overall order of importance structure on $[c_2]$, personnel data:

I1:	f_5	: employee history
I2:	f_6	: employee payroll history
I3:	f_{10}	: personnel pay/pos/rec changes
I4:	f_2, f_8, f_9, f_{11}	: personnel actions notices
I5:	f_1	: new employee data
I6:	f_{12}	: personnel grant application
I7:	f_3, f_{17}	: time reporting/correspondence

This is not necessarily an implementation order; an implementation order will be obtained via the P-analysis. This ordering might be termed a machineability index, or something similar. In this view we might say:

I1 & I2 : permanent global records

I3 - I5 : transactional records; facilitating and informational

I6 & I7 : informational; not machined

This order corroborates our intuitive sense of what is important among data sets.

The item (data set) which intuitively seems out of place is f_3 (employee time reporting), since employees here are salaried, and time and attendance reporting is largely pro forma; therefore we have little departmental interest beyond accounting. (As an aside, in the school system now this is a periodic report prepared on optical scan sheets.)

Looking at $[c_3]$, financial data, we find the same difficulty we had with $[c_1]$; neither U nor V provide us with a partition.

$$[c_3]\ U = [c_3] = \{f_{13}, f_{14}, f_{15}\}$$

$$[c_3]/V = [c_3] = \{f_{13}, f_{14}, f_{15}\}$$

Here, for each data set, the U-index is 5 and the V-index 3.5. Although we again have too small a subset of $[c_3]$ to work with, we might expect that the class of financial data may well be more cohesive in terms of scope of interest than some of the other classes; we note that our three sets from $[c_3]$ had a generally higher index of importance than either $[c_1]$ or $[c_2]$.

Since only one data set from $[c_7]$ is shown, no comment can be made about the class of system data.

Looking back to our ordering of personnel data, we need to comment briefly on disparities in size and content of data sets. At our level of definition we cannot say much about the content of the basic history records, which show the greatest importance. They are aggregate records, though, and if we exclude them from our list we should find that the transactional data provides at least a potential source of content information. An inspection of the transactional data sets in our example (I3-I5) corroborates this view.

SECTION III
RECOMMENDATIONS AND COMPUTATIONAL APPROACH

3.1 Plan of Action

In current practice the requirements definition and analysis phase is probably most arbitrary among the stages of development. It is uniformly recognized that formal analysis methods are highly dependent upon a complete and accurate definition of organizational functions and information requirements. That this kind of definition is not routinely obtainable is also recognized. Langefors, for example, as most workers do, acknowledges the difficulty and proceeds with his development of analytic method.[8] To highlight these difficulties, we have used a specific study (by no means a poor one) as source material in this paper. In this section, we will state the necessary content of such a report.

3.2 Derivation of Primary System Sets and Attributes

The great bulk of initial work is definitional in nature, comprising an inventory of a variety of structures. This can be seen to take place naturally in four stages:

3.2.1 Derivation of D, G, M, and C

Specification of these four sets provides the basic skeleton upon which the system is constructed. The set D is the most available and apparent. Its specification is a recording function in which notational devices are added for convenience. In many organizations some such coded notation is already available; this is especially true of the Federal government and military establishment.

The set of object classes C is usually intuitively obtainable, with two exceptions. Ambiguities occasionally arise because of confusion between the object of discussion and the intended uses of information related to the object. A man's pay and payroll funds must be distinguished as descriptive of different object classes - i.e. personnel and

8. See Langefors, Vol. II, Chapter 5, pp. 224-225.

funds. In addition, there exists that set of information which can be said to pertain only to the system as a whole; for example, authorized manpower structure, funding authority, or external environment. These must be distinguished from the object classes which they influence.

The definition of the sets G and M are the most difficult, since they require a high-level understanding of how the organization functions and should function. Senior management must therefore be responsible for defining these sets and establishing internal priorities. G should closely match the organizational objectives, while M should reflect the operational functioning of the organization.

3.2.2 Derivation of the (S, I, O)-structure

Using the D-structure as a checklist, it is necessary to define, by department, the elements of S, I, and O and to assign secondary attributes. For each p_i, the program, activity, predecessor function and successor function must be identified; for each i_i, the object class, programs, and activities must be identified. (Note that multiple assignments of program and activity to a data element is to be expected.) Similarly, to each o_i, frequency, activities, and source/processed must be assigned.

This is a major cataloging activity; and many of the deficiencies in requirements-analysis stem from the inability to carry this out adequately. For success, two factors are important. First, derivation must be conducted uniformly and at a feasible level of detail; greater detail is acceptable only if it requires no additional expenditure in collection or in processing. Second, precise uniformity in descriptive terminology is required so that identical or related elements may be recognized.

3.2.3 Derivation of P, F, and R

Once the work of the first two steps have been completed, use of the computer may be considered. We wish to generate P from the sets S_i and we wish to generate F from the input sets I_i; if the work in step 2 above was successful, this is a straightforward machine task. Necessary consolidations and sorting can be performed efficiently and quickly by

computer. As byproducts of computer processing we can identify ambiguities, overlaps (a problem in manual consolidation), and omissions due to incomplete collection. In addition, computer generation of these sets allows us to use the machine again in subsequent steps.

3.2.4 Generation of W and Indices on F and R

Working from the predecessor and successor data on each element of P, the computer provides an efficient generator for what is a most tedious hand-operation--the generation of the set of processes W. This is a two-step process: the computer is used first to build strings representing candidate processes, and then to eliminate redundancies and to consolidate. (Some manual editing may be required to reflect discontinuities not visible from the data itself; these revisions should be incorporated into W before proceeding).

Working from the attributed values of G and M, the computer can be used to calculate values of U and V for the elements of F, and values of X for the elements of R. Values for U and V are then added to the individual f_i records, and values for X to the corresponding r_i records.

3.3 Procedures for Preliminary Analysis

The developments in Sections 2.2 and 2.3 were experimental and demonstrative in nature, intended to suggest ways in which specific guidance to the implementation of an information system could be derived from the requirements data alone. The methods, however, need further use over more complete data before being formalized. In this section, then, we can merely suggest some general procedural approaches to the analysis of requirements data.

3.3.1 Analysis of P

In Section 2.2 we generated successive partitions (refinements) of P utilizing the induced attributes from G, M, and W, and producing an order on the subsets of functions so obtained. The procedure is:

1. Partition P according to G and associate with the subsets generated the priority of the corresponding g_i

2. Partition the elements of P/G according to M.

3. Generate a precedence-order based on W for each element of P/G/M which is populous, i.e., which has more than one function within it. Use graph-theoretic methods to establish internal priorities.

4. Use priorities of M or counting methods for breaking ties, or for further ordering where necessary.

Because of natural ties and ambiguities which may develop in the use of the W-structure, it may be advantageous to use an interactive method, in which the computer is used for clerical work and decisions are reserved for a human analyst. In any event, further experimentation is needed before more precise procedures can be stated.

3.3.2 Analysis of F

In Section 2.3 we considered each object class of data separately and used the computed index values to derive an order of importance among the data sets. Since some of the critical decisions which were made are not readily formulated, we can only outline an approximate procedure:

1. Sort the set F by object class and inspect each separately.

2. Sort data sets within an object class and group them according to decreasing U-index. Initial assignment of importance is according to highest U-value down.

3. Where more than one data set occupies a U-value, the group is again ordered, this time by decreasing V-index. Since values of V are not necessarily discrete, it is necessary to impose arbitrary half-closed intervals prior to ordering. This, however, should cause no difficulty in processing. Overall importance-ranks are then reassigned incorporating this secondary order.

4. The list is scanned top-down, removing those data sets which are permanent global files. This requires human judgment; however, once identified as such, this should become an element of identification for the set.

5. The list is scanned bottom-up, removing all data sets with importance below a minimal UV-value. These items will be culled manually to insure that important data sets are not discarded, but generally these will be items which should not be considered in automation.

The resultant list, possibly renumbered, provides guidance as to those data sets which must be incorporated into global files and indicates an order of importance for machine conversion.

3.4 Summary

In this paper an attempt has been made to contribute a level of formality to the requirements phase of system development and to the system-level aspects of the subsequent analysis. The intent has been to replace much of the intuitive approach to implementation with analytic method. Methods have been displayed which generate concrete recommendations to implementation planning. Finally, the use of the computer as an aid to analysis of requirements has been suggested.

APPENDIX
LITERATURE NOTES

Beyond the restricted methods of traditional systems-and-procedures workers, little has been written concerning formal recording and analysis methods for the requirements phase. In the business systems area, a few efforts of minor interest can be noted. Grindley (3,4,13) in Great Britain has suggested a recording method he calls "Systematics" which provides a structure that builds from object classes. Sussams (12), also of Great Britain, suggests a charting method which records "operators" (programs, specifications), "operands" (material, data), and "operatives" (men, machines) in an attempt to deal with the complexity of information system design. In this country, Stevens (11) suggests the use of a machined file for recording an even more detailed file on data elements than is suggested in this paper.

Granting that the essential recording -- in particular the recording of the (S, I, O) structure -- has been accomplished, a number of substantial works deal with the analysis of this material. Langefors (7), Salton (9,10), and Zunde (14) attack the analysis of information systems with linear algebra and graph theory. Graph-theoretic methods provide a powerful handle for the processing of structured data. Two works, those of Harary et. al. (5) and Busacker and Saaty (1), provide a broad entry into this material. Finally, two papers are included in the bibliography which are suggestive of ways in which graph-theoretic methods can provide excellent results. In Luckman's paper (8), the use of option graphs eliminates difficulties of interdependence in design decisions. Carroll (2) uses graph methods to analyse activity within an existing system.

BIBLIOGRAPHY

1. Busacker, R. G., and Saaty, T. L. Finite Graphs and Networks. New York, McGraw-Hill, 1965.

2. Carroll, J. M. "Methodology for Information System Analysis," J. Ind. Eng., V 18:11 (Nov 1967), 650-657.

3. Grindley, C. B. B. "Systematics; A Non-programming Language for Designing and Specifying Commercial Systems for Computers," Computer J. 9 (Aug 1966), 124-128.

4. Grindley, C. B. B. "Use of Decision Tables Within Systematics," Computer J. 11 (Aug 1968), 128-133.

5. Harary, Frank, Norman, R. Z., and Cartwright, D. Structural Models: An Introduction to the Theory of Directed Graphs. New York, John Wiley and Sons, 1968.

6. IBM Federal Systems Division. User requirements report for the Atlanta Public Schools Education Information System, Atlanta. (October 14, 1968).

7. Langefors, Borje. Theoretical Analysis of Information Systems, Vols. I and II. Lund, Sweden, Studentlitteratur, 1966.

8. Luckman, J. "Approach to Management of Design," Operational Research Q. 18:4 (Dec 1967), 345-358.

9. Salton, Gerard. Automatic Information Organization and Retrieval. New York, McGraw-Hill, 1968.

10. Salton, Gerard. "Information Dissemination and Automatic Information Systems," Proc. IEEE 54:12 (Dec 1966), 1663-1678.

11. Stevens, W. B. "The Concept of the Data Analysis and Control Catalog for MIS," Comp. and Autom. 17 (April 1968), 40-42.

12. Sussams, J. E. "Business Systems Analysis," Operational Research Q. 19: Symp. (April 1968), 85-90.

13. Tatham, Laura. "A 'Systematic' Approach to Systems Analysis," Data Proc. 9:1 (Jan-Feb 1967) 18-21.

14. Zunde, Pranas. Course notes for a course in information systems design, Georgia Tech, Summer 1968, unpublished.